

Using Volunteer Computing for Mounting SAT-based Cryptographic Attacks

Alexander Semenov, Oleg Zaikin, and Ilya Otpuschennikov

Institute for System Dynamics and Control Theory SB RAS, Irkutsk, Russia
biclop Rambler@yandex.ru, zaikin.icc@gmail.com, otilya@yandex.ru

Abstract. In this paper we describe the volunteer computing project SAT@home, developed and maintained by us. This project is aimed at solving hard instances of the Boolean satisfiability problem (SAT). We believe that this project can be a useful tool for computational study of inversion problems of some cryptographic functions. In particular we describe a series of experiments performed in SAT@home on the cryptanalysis of the widely known keystream generator A5/1. In all experiments we analyzed one known burst (114 bits) of keystream produced by A5/1. Before the cryptanalysis itself there is a stage on which the partitioning of the original problem to a family of subproblems is carried out. Each of subproblems should be easy enough so that it could be solved in relatively small amount of time by volunteer's PC. We construct such partitioning using the special technique based on the Monte Carlo method and discrete optimization algorithms for special predictive functions. Besides this in the paper we describe the technique for reducing inversion problems of cryptographic functions to SAT.

Keywords: Boolean satisfiability problem, SAT, volunteer computing, BOINC, cryptanalysis, A5/1, SAT@home

1 Introduction

In this paper we present our experience of the development of the volunteer computing project SAT@home aimed at solving hard instances of the Boolean satisfiability problem (SAT). In particular this project can be used to justify from the practical point of view the resistance of some cryptographic functions to attacks based on the algorithms for solving SAT. We confirm this assertion by successfully solving in SAT@home the series of cryptanalysis problems of widely known keystream generator A5/1, based on one known burst (114 bits) of keystream.

The Boolean satisfiability problem is formulated as follows: for an arbitrary Boolean formula to determine if there exists such assignment of its variables, that makes this formula true. This problem is NP-complete, and if $P \neq NP$ there are no polynomial algorithms that could solve it. On the other hand, the class of practically important problems that can be effectively reduced to SAT is very wide [3]. Because of this fact it is important to develop the algorithms that show good effectiveness in application to various cases of SAT.

In practice it is usually convenient to consider satisfiability problem for a formula presented in the Conjunctive Normal Form (CNF). Using the Tseitin transformations [41] it is possible to effectively (in polynomial time) reduce the SAT for an arbitrary Boolean formula to the SAT for some CNF.

In the last 10 years the effectiveness of SAT solving algorithms has significantly increased. Today SAT approach is used to solve various problems from the areas of symbolic verification, bioinformatics, discrete optimization, combinatorial design, cryptanalysis, etc. There are several basic classes of algorithms used to construct modern SAT solvers. All the results presented in this paper were obtained using the so called Conflict Driven Clause Learning (CDCL) solvers, the description of main architectural principles of which can be found, for example, in [28].

To solve SAT instances encoding the problems of cryptanalysis of ciphering systems used in the real world it is insufficient to employ only the resources of a single PC. In this context the development of SAT solving methods, designed for parallel and distributed computing environments, is highly relevant. There are several general concepts of solving SAT in parallel. From our point of view for solving SAT instances, encoding cryptographic attacks, the volunteer computing concept looks promising, because this kind of distributed computing makes it possible to solve one particular computing problem for a prolonged time with relatively high performance (on average).

It should be noted, that when solving a particular problem in a volunteer computing project there apply some restrictions. First, the problem should be divided into subproblems in such a way that each subproblem can be solved by a volunteer's PC in relatively small amount of time (several hours at most). Also it is worth mentioning that volunteers PCs do not interconnect directly with each other. From these restrictions it is natural to draw the conclusion that to parallelize the original problem it is best to use the partitioning approach. The partitioning method must make it possible to estimate the time required to solve all produced subproblems. Below we describe such method and apply it to the parallelization of cryptanalysis problems for the widely known keystream generator A5/1. Then we process the partitionings obtained in parallel using the volunteer computing project SAT@home.

Let us give a brief outline of the paper. In the second section we present the general technique used to reduce cryptographic problems to SAT and describe the TRANSALG program system implementing this technique. In the third section we give brief description of main SAT solving algorithms that we applied to cryptographic problems. In the fourth section we propose an approach to the parallelization of SAT instances that encode cryptanalysis problems. Thus, we describe the general technique of constructing SAT partitionings that can be processed in distributed computing environments. In the fifth section we present the results of the computational experiments on the search for SAT partitionings, with the use of the algorithms described in the previous section. In the sixth section we describe the volunteer computing project SAT@home developed by us specifically for the purpose of solving hard SAT instances, and present the results of computational experiments held in SAT@home on the cryptanalysis of the A5/1 keystream generator.

2 Reducing Inversion Problems of Cryptographic Functions to SAT

Hereinafter by $\{0, 1\}^l$ we denote the set formed by all possible binary words of length l . By $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ we denote a discrete function that transforms binary words of length n to binary words of length m . For each function considered below we assume that this function is defined everywhere on $\{0, 1\}^n$ and is specified by some algorithm A . By $\text{Range } f$, $\text{Range } f \subseteq \{0, 1\}^m$ we denote the set of all output values of the function f . Let us consider the following problem: for a given $y \in \text{Range } f$ and the known algorithm A we need to find such $x \in \{0, 1\}^n$, that $f(x) = y$. We will refer to this general problem as to the problem of inversion of function f . Many cryptanalysis problems can be considered in its context. For example, let f be a function specified by the algorithm of some keystream generator. Then $x \in \{0, 1\}^n$ is the secret key and $y \in \{0, 1\}^m : f(x) = y$ is the keystream produced by the generator considered. In the vast majority of keystream ciphers the ciphertext is formed by adding keystream and some secret message M in arithmetics modulo 2. Sometimes, it is possible (usually when the protocol is inconsistent) that some message M becomes known. It may represent some technical data and thus not even contain any valuable information (as in case of the ciphering protocol used in the GSM standard). However, if we know M and the corresponding ciphertext then we can find y and consider the problem of search for the secret key $x : f(x) = y$. Knowing x makes it possible to read any message ciphered using this secret key. The problem of determining x based on a given y and known algorithm specifying function f is usually called a known plaintext attack. Hereinafter we will study this particular cryptographic problem.

2.1 Theoretical Basis of SAT Encoding of Discrete Functions Inversion Problems

Let us describe the general scheme of encoding the problem of inversion of f to SAT. The main idea that underlies such reductions is based on the Cook theorem [8]. It consists in forming a correspondence between the computation of function f by algorithm A and some system of Boolean equations. The complexity of the procedure that constructs the system of Boolean equations is bounded above by a polynomial of the complexity of algorithm A .

In more detail the transition from the problem of inversion of f to the CNF encoding looks as follows. By $S(f)$ we denote the Boolean Circuit over the basis $\{\wedge, \neg\}$ that implements function f . In the circuit $S(f)$ there are n inputs and m outputs. Let us link Boolean variables x_1, \dots, x_n , forming the set X , with inputs of the circuit, and Boolean variables y_1, \dots, y_m , forming the set Y , with its outputs. We also mark each logical gate G by an auxiliary variable $v(G)$ such that $v(G) \notin X$. The set of all auxiliary variables we denote by V . Note that $Y \subseteq V$. Let $v(G)$ be an arbitrary variable from V and G be a corresponding gate. If G is a NOT-gate and $u \in X \cup V$ is the variable linked with the input of G then we encode the gate G with the Boolean formula $v(G) \equiv \neg u$ (hereinafter by \equiv we mean logical equivalence). If G is an AND-gate and $u, w \in X \cup V$ are variables linked with the gate inputs then we encode G with the formula $v(G) \equiv u \wedge w$. CNF-representations of Boolean functions specified by formulas $v(G) \equiv$

$\neg u$ and $v(G) \equiv u \wedge v$ look as follows (respectively):

$$\begin{aligned} & (v(G) \vee u) \wedge (\neg v(G) \vee \neg u) \\ & (v(G) \vee \neg u \vee \neg w) \wedge (\neg v(G) \vee u) \wedge (\neg v(G) \vee w) \end{aligned} \quad (1)$$

Thus with an arbitrary gate G of circuit $S(f)$ we associate a CNF $C(G)$ of the kind (1). We will say that the CNF $C(G)$ encodes the gate G , and the CNF

$$C(f) = \bigwedge_{G \in S(f)} C(G)$$

encodes the circuit $S(f)$. The described technique of constructing a CNF for a circuit $S(f)$ is known as Tseitin transformations [41].

Now let $(\beta_1, \dots, \beta_m) \in \{0, 1\}^m$ be an arbitrary assignment of variables y_1, \dots, y_m . Consider the CNF

$$C(f) \wedge y_1^{\beta_1} \wedge \dots \wedge y_m^{\beta_m} \quad (2)$$

where

$$y^\beta = \begin{cases} \neg y, & \text{if } \beta = 0 \\ y, & \text{if } \beta = 1 \end{cases}$$

From the properties of the Tseitin transformations it follows that if $(\beta_1, \dots, \beta_m) \in \text{Range } f$ then the CNF (2) is satisfiable and from any of its satisfying assignments one can effectively extract such assignment $x_1 = \alpha_1, \dots, x_n = \alpha_n$, $\alpha_i \in \{0, 1\}$, $i \in \{1, \dots, n\}$, that $f(\alpha_1, \dots, \alpha_n) = (\beta_1, \dots, \beta_m)$.

Thus, using the technique described above we reduce the problem of inversion of function f in an arbitrary point $y \in \text{Range } f$ to SAT for the CNF (2). It makes it possible to apply the SAT solvers to the latter problem.

2.2 Encoding Inversion Problems of Cryptographic Functions to SAT Using the Transalg System

In this section we will briefly describe the TRANSALG program system [34] that was designed specifically to reduce inversion problems of cryptographic functions to SAT. The original function f should first be represented in the form of the special TA-program A . TA-programs have C-like syntax. The translation of an arbitrary TA-program A essentially involves the symbolic execution of A [26], and results in the propositional encoding of A in the form of CNF. In other words the result of compilation of the program A is not a machine code but a set of Boolean formulas constructed according to the Tseitin transformations. In more detail, the result of the compilation of the TA-program is the CNF $C(f)$ encoding the circuit $S(f)$. By fixing the values of output variables $(\beta_1, \dots, \beta_m)$ (essentially transitioning to CNF (2)) we obtain a SAT instance that encodes the inversion problem of function f in the point $y = (\beta_1, \dots, \beta_m)$.

Below the main object of our study is the A5/1 keystream generator that has been used to encrypt data within the GSM standard for more than 15 years. On Figure 1 we demonstrate the scheme of the A5/1 generator from the paper [4].

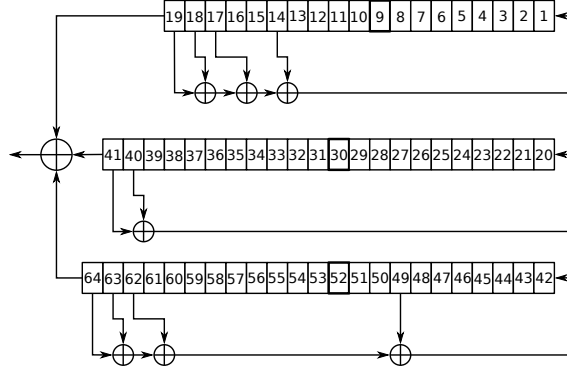


Fig. 1. The A5/1 keystream generator

The generator contains three linear feedback shift registers (LFSRs) specified by the following feedback polynomials:

$$\begin{aligned} LFSR1 & X^{19} + X^{18} + X^{17} + X^{14} + 1; \\ LFSR2 & X^{22} + X^{21} + 1; \\ LFSR3 & X^{23} + X^{22} + X^{21} + X^8 + 1. \end{aligned}$$

The registers in A5/1 are shifted asynchronously. Assume that at the initial time moment $t = 0$ the registers contain the secret key. By b_i^t , $i \in \{1, 2, 3\}$ denote the value of the clocking bit of the i -th register at time moment t . The clocking bits are contained in cells 9, 30 and 52. With each clocking bit the special function

$$\tau_i(t) = \begin{cases} 1, & \text{if } \text{majority}(b_1(t), b_2(t), b_3(t)) = b_i(t) \\ 0, & \text{if } \text{majority}(b_1(t), b_2(t), b_3(t)) \neq b_i(t) \end{cases}$$

is associated. Here $\text{majority}(A, B, C) = (A \wedge B) \vee (A \wedge C) \vee (B \wedge C)$. The register number $i \in \{1, 2, 3\}$ is shifted at time moment $t \in \{1, 2, \dots\}$ only if $\tau_i(t) = 1$. Otherwise the register retains the condition in which it was at the previous time moment. At each time moment the generator outputs one bit of keystream that is produced as a sum modulo 2 of bits contained in cells 19, 41 and 64. The length of the secret key in the A5/1 keystream generator is 64 bits.

On Figure 2 we present the TA-program specifying the discrete function

$$f_{A5/1} : \{0, 1\}^{64} \rightarrow \{0, 1\}^{114}. \quad (3)$$

This function generates 114 bits of keystream from the secret key of length 64 bits according to the A5/1 algorithm. The length of keystream fragment is chosen to be equal to 114 bits because in the GSM standard the information is encrypted by blocks of 114 bits. Such blocks are called bursts.

The result of the compilation of the program from the Figure 2 is the CNF with 7816 variables and 35568 clauses (size of the file in DIMACS format is 730 kilobytes).

```

define len 114
__in bit regA[19];
__in bit regB[22];
__in bit regC[23];
__out bit stream[len];
int midA = 8;
int midB = 10;
int midC = 10;

bit majority(bit x, bit y, bit z){
    return x & y | x & z | y & z;
}

void main(){
    for(int i = 0; i < len; i = i + 1){
        bit maj = majority(regA[midA], regB[midB], regC[midC]);
        if(!(maj^regA[midA])) shift_regA();
        if(!(maj^regB[midB])) shift_regB();
        if(!(maj^regC[midC])) shift_regC();
        stream[i] = regA[18]^regB[21]^regC[22];
    }
}

```

Fig. 2. TA-program specifying function $f_{A5/1}$.

3 Algorithms for Solving SAT Instances that Encode Inversion Problems of Cryptographic Functions

Main classes of the SAT solving algorithms were extensively described in [3]. After many computational experiments we made the conclusion that the algorithms based on the concept of Conflict Driven Clause Learning (CDCL) [28] perform best on cryptographic problems. All the modern CDCL solvers that show good effectiveness on wide classes of industrial tests are based on the DPLL algorithm [9]. However, in its pure form the DPLL algorithm has relatively low performance. The main idea of the CDCL concept is to complement the DPLL with an ability to use memory to store the information about traversed parts of the search tree. This information is stored in the form of restrictive constraints which are called conflict clauses. The solver GRASP [29] was the first full-fledged CDCL solver. In the subsequent papers [45,33] several important algorithmical constructions were described that made it possible to improve the CDCL concept. It led to the creation of SAT solvers widely applied nowadays to combinatorial problems from various areas.

Active development of parallel SAT solving algorithms started quite recently. First parallel SAT solvers competition was held in 2008. There are two main approaches to parallel SAT solving: the partitioning approach and the portfolio approach. In the partitioning approach the search space is divided into separate subspaces that are processed in parallel. In the portfolio approach all computing processes work with the same original search space, but they use different initial parameters and search heuristics, and also can share the collected data with each other.

The SAT instances encoding the inversion problems of cryptographic functions used in the real world usually are very hard, and to solve them it is insufficient to use the

resources of a usual PC. The graphical processors (GPU) in the context of solving SAT nowadays do not yield any advantages in comparison with traditional CPUs. That is explained by the fact that modern GPUs have relatively high memory latencies, which significantly decrease the performance of CDCL algorithms.

For solving hard SAT instances one can use computing clusters with nodes based on CPUs. The portfolio approach makes it impossible to effectively utilize a large number of cluster nodes because intensive conflict clauses exchange greatly taxes the inter-connection environment. The advantage of the partitioning approach is that it doesn't limit the amount of computing nodes of the cluster that can be used. The low intensity of data exchange between computing processes in this case makes it possible to employ not only computing clusters but also distributed computing systems. The use of such systems (including volunteer computing projects) suits better for solving hard SAT instances because of the low computing costs (in comparison to that of clusters). To organize the computations in the distributed system the original problem should be divided into a family of significantly less complex subproblems in such a way that each subproblem can be solved by a computing node (such as, for example, a usual PC) in relatively small amount of time. In the next section we will describe the SAT partitioning technique that we used to organize the computations in grid systems.

4 Partitioning Procedures for SAT Instances Encoding Inversion Problems of Cryptographic Functions

So as we explained above we use the partitioning approach to parallelize the SAT solving process. There exist many various ways of constructing SAT partitionings [19]. Moreover, even following one way it is possible to make several different partitionings for the same SAT instance. The main problem that arises in this context is to justify why one partitioning is better than another. The basic partitioning effectiveness criterion is the time required to process it in a distributed computing environment. Unfortunately, not for all types of SAT partitionings it is possible to evaluate their effectiveness. In this section we will describe one type of SAT partitioning for which there exists quite natural procedure for constructing estimations of time required to process it. We used this type of partitioning to solve inversion problems of some cryptographic functions in parallel.

Assume that C is the CNF encoding the inversion problem of some discrete function. By X denote the set of Boolean variables that appear in CNF C . By \tilde{X} denote an arbitrary subset of X . Let α be an arbitrary set of values of variables from \tilde{X} . By $C[\tilde{X}/\alpha]$ we denote the CNF produced from C by setting values from α to corresponding variables. Let us consider the set $\Delta(C, \tilde{X}) = \{C[\tilde{X}/\alpha]\}_{\alpha \in \{0,1\}^{|\tilde{X}|}}$. It is easy to see that the CNF C is satisfiable if and only if at least one CNF from the set $\Delta(C, \tilde{X})$ is satisfiable. Therefore, the set $\Delta(C, \tilde{X})$ is a partitioning of the SAT instance C [19]. We call the set \tilde{X} that produces the partitioning $\Delta(C, \tilde{X})$ a *decomposition set*.

Now let us describe the approach to constructing time estimations for the processing of SAT partitionings that was proposed in [37]. First let us fix some particular SAT solving algorithm A . Hereinafter we assume that A is a complete algorithm, i.e. it has finite runtime for an arbitrary input. Now consider an arbitrary CNF C over the set X of Boolean variables. Let \tilde{X} be an arbitrary decomposition set. Consider the set $\{0, 1\}^{|\tilde{X}|}$. Let us define the uniform distribution on this set. With a randomly selected vector $\alpha \in \{0, 1\}^{|\tilde{X}|}$ we associate the runtime of algorithm A given the CNF $C \left[\tilde{X}/\alpha \right]$ as an input. This time can be considered as a value of some random variable which we denote below as $\xi_A \left(C, \tilde{X} \right)$. Since A is a complete algorithm, then the random variable $\xi_A \left(C, \tilde{X} \right)$ takes only finite values with some probabilities. Thus, the expected value of $\xi_A \left(C, \tilde{X} \right)$ is finite. By $t_A \left(C, \tilde{X} \right)$ denote the time required to process the SAT partitioning $\Delta \left(C, \tilde{X} \right)$ sequentially by algorithm A . In [37] it was shown that under listed conditions the following equality holds:

$$t_A \left(C, \tilde{X} \right) = 2^{|\tilde{X}|} \cdot \mathbb{E} \left[\xi_A \left(C, \tilde{X} \right) \right] \quad (4)$$

To estimate the expected value $\mathbb{E} \left[\xi_A \left(C, \tilde{X} \right) \right]$ it is possible to use the computational scheme of the Monte Carlo method in its classical form [32]. Namely, for an arbitrary decomposition set \tilde{X} we randomly select vectors $\alpha_1, \dots, \alpha_N$ from $\{0, 1\}^{|\tilde{X}|}$. Then for each $\alpha_j, j \in \{1, \dots, N\}$ we measure the time required by algorithm A to solve the SAT for CNF $C \left[\tilde{X}/\alpha_j \right]$. We denote the time obtained by ξ^j . After this we compute the value $\frac{1}{N} \cdot \sum_{j=1}^N \xi^j$. In accordance with the theoretical basis of the Monte Carlo method [32,24], when N is large enough, the value $\frac{1}{N} \cdot \sum_{j=1}^N \xi^j$ becomes a good estimation of the expected value $\mathbb{E} \left[\xi_A \left(C, \tilde{X} \right) \right]$. Therefore it follows from (4) that the value

$$F_A \left(C, \tilde{X} \right) = \frac{2^{|\tilde{X}|}}{N} \cdot \sum_{j=1}^N \xi^j \quad (5)$$

is the estimation of time required by algorithm A to sequentially process the whole SAT partitioning $\Delta \left(C, \tilde{X} \right)$ and this estimation is the more accurate the larger N is. Below we will refer to the function defined according to (5) as to the *predictive function*.

As we already noted above, different partitionings of the same SAT problem can have different values of $t_A \left(C, \tilde{X} \right)$. In practice it is important to be able to find partitionings that can be processed in affordable time. Below we will describe a scheme of automatic search for good partitionings that is based on the procedure minimizing the predictive function value in a special search space.

So we consider the satisfiability problem for some CNF C . Let $X = \{x_1, \dots, x_M\}$ be a set of all Boolean variables in this CNF and $\tilde{X} \subseteq X$ be an arbitrary decomposition

set. The set \tilde{X} can be represented by the binary vector $\chi = (\chi_1, \dots, \chi_M)$. Here

$$\chi_i = \begin{cases} 1, & \text{if } x_i \in \tilde{X} \\ 0, & \text{if } x_i \notin \tilde{X} \end{cases}, i \in \{1, \dots, M\}$$

With an arbitrary vector $\chi \in \{0, 1\}^M$ we associate the value of function $F(\chi)$ computed in the following manner. For vector χ we construct the corresponding set \tilde{X} (it is formed by variables from X that correspond to 1 positions in χ). Then we generate a random sample $\{\alpha_1, \dots, \alpha_N\}$, $\alpha_j \in \{0, 1\}^{|\tilde{X}|}$ and solve SAT for CNFs $C[\tilde{X}/\alpha_j]$. For each of these SAT instances we measure ξ^j — the runtime of algorithm A on the input $C[\tilde{X}/\alpha_j]$. After this we calculate the value of $F_A(C, \tilde{X})$ according to (5). As a result we have the value of $F(\chi)$ in the considered point of the search space.

Now we will solve the problem $F(\chi) \rightarrow \min$ over the set $\{0, 1\}^M$. Of course, the problem of search for the exact minimum of function $F(\chi)$ is extraordinarily complex. Therefore our main goal is to find in affordable time the points in $\{0, 1\}^M$ with relatively good values of function $F(\cdot)$. For this purpose we propose two minimization algorithms. The first employs the simulated annealing algorithm, while the second is based on the tabu search concept. We will compare the effectiveness of these algorithms on the example of the problem of search for the good SAT partitioning for the SAT instances encoding the cryptanalysis of the A5/1 keystream generator (i.e. the inversion problem of (3)).

First we need to introduce the notation. By \mathfrak{R} we denote the search space, for example, $\mathfrak{R} = \{0, 1\}^M$, however, as we will see later, for the problems considered one can use the search spaces of much less power. The minimization of function $F(\cdot)$ is considered as an iterative process of transitioning between the points of the search space:

$$\chi^0 \rightarrow \chi^1 \rightarrow \dots \rightarrow \chi^i \rightarrow \dots \rightarrow \chi^*.$$

By $N_\rho(\chi)$ we denote the neighborhood of point χ of radius ρ in the search space \mathfrak{R} . The point from which the search starts we denote as χ_{start} . The current Best Known Value of $F(\cdot)$ is denoted by F_{best} . The point in which the F_{best} was achieved we denote as χ_{best} . By χ_{center} we denote the point the neighborhood of which is processed at the current moment. We call the point, in which we computed the value $F(\cdot)$ a *checked point*. The neighborhood $N_\rho(\chi)$ in which all the points are checked is called *checked neighborhood*. Otherwise the neighborhood is called *unchecked*.

According to the scheme of the simulated annealing [27], the transition from χ^i to χ^{i+1} is performed in two stages. First we choose a point $\tilde{\chi}^i$ from $N_\rho(\chi^i)$. The point $\tilde{\chi}^i$ becomes the point χ^{i+1} with the probability denoted as $\Pr\{\tilde{\chi}^i \rightarrow \chi^{i+1} | \chi^i\}$. This probability is defined in the following way:

$$\Pr\{\tilde{\chi}^i \rightarrow \chi^{i+1} | \chi^i\} = \begin{cases} 1, & \text{if } F(\tilde{\chi}^i) < F(\chi^i) \\ \exp\left(-\frac{F(\tilde{\chi}^i) - F(\chi^i)}{T_i}\right), & \text{if } F(\tilde{\chi}^i) \geq F(\chi^i) \end{cases}$$

In the pseudocode of the algorithm demonstrated below the function that tests if the point $\tilde{\chi}^i$ becomes χ^{i+1} , is called `PointAccepted` (this function returns the value of

true if the transition occurs and false otherwise). The change of parameter T_i corresponds to decreasing the “temperature of the environment” [27] (in the pseudocode by `decreaseTemperature()` we denote the function which implements this procedure). Usually it is assumed that $T_i = Q \cdot T_{i-1}$, $i \geq 1$, where $Q \in (0, 1)$. The process starts at some initial value T_0 and continues until the temperature drops below some threshold value T_{inf} (in the pseudocode the function that checks this condition is called `temperatureLimitReached()`).

Algorithm 1: Simulated annealing algorithm for minimization of the predictive function

Input: CNF C , initial point χ_{start}
Output: Pair $\langle \chi_{best}, F_{best} \rangle$, where F_{best} is a prediction for C , χ_{best} is a corresponding decomposition set

```

1  $\langle \chi_{center}, F_{best} \rangle \leftarrow \langle \chi_{start}, F(\chi_{start}) \rangle$ 
2 repeat
3    $bestValueUpdated \leftarrow false$ 
4    $\rho = 1$ 
5   repeat // check neighborhood
6      $\chi \leftarrow$  any unchecked point from  $N_\rho(\chi_{center})$ 
7     compute  $F(\chi)$ 
8     mark  $\chi$  as checked point in  $N_\rho(\chi_{center})$ 
9     if PointAccepted( $\chi$ ) then
10       $\langle \chi_{best}, F_{best} \rangle \leftarrow \langle \chi, F(\chi) \rangle$ 
11       $\chi_{center} \leftarrow \chi_{best}$ 
12       $bestValueUpdated \leftarrow true$ 
13     if ( $N_\rho(\chi_{center})$  is checked) and (not  $bestValueUpdated$ ) then
14        $\rho = \rho + 1$ 
15     decreaseTemperature()
16   until  $bestValueUpdated$ 
17 until timeExceeded() or temperatureLimitReached()
18 return  $\langle \chi_{best}, F_{best} \rangle$ 

```

In the other approach to the minimization of $F(\cdot)$ we employed the tabu search scheme [14]. According to this approach the points from the search space, in which we already calculated the values of function $F(\cdot)$ are stored in special tabu lists. When we try to improve the current Best Known Value of $F(\cdot)$ in the neighborhood of some point χ_{center} then for an arbitrary point χ from the neighborhood we first check if we haven't computed $F(\chi)$ earlier. If we haven't and, therefore, the point χ is not contained in tabu lists, then we compute $F(\chi)$. This strategy is justified in the case of the minimization of predictive function $F(\cdot)$ because the computing of values of the function in some points of the search space is very expensive. The use of tabu lists makes it possible to significantly increase the number of points of the search space processed per time unit.

Let us describe the Tabu Search algorithm (TS-algorithm) for minimization $F(\cdot)$ in more detail. To store the information about points, in which we already computed the

value of $F(\cdot)$ we use two tabu lists L_1 and L_2 . The L_1 list contains only points with checked neighborhoods. The L_2 list contains checked points with unchecked neighborhoods. Below we present the pseudocode of the TS-algorithm for $F(\cdot)$ minimization.

Algorithm 2: Tabu search algorithm for minimization of the predictive function

Input: CNF C , initial point χ_{start}
Output: Pair $\langle \chi_{best}, F_{best} \rangle$, where F_{best} is a prediction for C , χ_{best} is a corresponding decomposition set

```

1  $\langle \chi_{center}, F_{best} \rangle \leftarrow \langle \chi_{start}, F(\chi_{start}) \rangle$ 
2  $\langle L_1, L_2 \rangle \leftarrow \langle \emptyset, \chi_{start} \rangle$  // initialize tabu lists
3 repeat
4    $bestValueUpdated \leftarrow false$ 
5   repeat // check neighborhood
6      $\chi \leftarrow$  any unchecked point from  $N_\rho(\chi_{center})$ 
7     compute  $F(\chi)$ 
8      $markPointInTabuLists(\chi, L_1, L_2)$  // update tabu lists
9     if  $F(\chi) < F_{best}$  then
10        $\langle \chi_{best}, F_{best} \rangle \leftarrow \langle \chi, F(\chi) \rangle$ 
11        $bestValueUpdated \leftarrow true$ 
12   until  $N_\rho(\chi_{center})$  is checked
13   if  $bestValueUpdated$  then  $\chi_{center} \leftarrow \chi_{best}$ 
14   else  $\chi_{center} \leftarrow getPoint(L_2)$ 
15 until  $timeExceeded()$  or  $L_2 = \emptyset$ 
16 return  $\langle \chi_{best}, F_{best} \rangle$ 

```

In this algorithm the function $markPointInTabuLists(\chi, L_1, L_2)$ adds the point χ to L_2 and then marks χ as checked in all neighborhoods of points from L_2 that contain χ . If as a result the neighborhood of some point χ' becomes checked, the point χ' is removed from L_2 and is added to L_1 . If we have processed all the points in the neighborhood of χ_{center} but could not improve the F_{best} then as the new point χ_{center} we choose some point from L_2 . It is done via the function $getPoint(L_2)$. To choose the new point in this case one can use various heuristics. At the moment the TS-algorithm employs the following heuristics: it chooses the point for which the total conflict activity of Boolean variables, contained in the corresponding decomposition set, is the largest.

As we already mentioned above, taking into account the features of the considered SAT problems makes it possible to significantly decrease the size of the search space. For example, knowing the so called Backdoor Sets [42] can help in that matter. In case of the SAT instances encoding the inversion problems of functions of the kind $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ the set \tilde{X}_{in} formed by the variables encoding the inputs of the Boolean circuit $S(f)$ is the so called Strong Unit Propagation Backdoor Set [22]. It means that if we use \tilde{X}_{in} as a decomposition set, then the CDCL solver will solve SAT for any CNF of the kind $C \left[\tilde{X}_{in}/\alpha \right]$, $\alpha \in \{0, 1\}^{|\tilde{X}_{in}|}$ on the preprocessing stage,

i.e. very fast. Therefore the point χ_{in} that corresponds to the set \tilde{X}_{in} can be used as an initial point in the predictive function minimization procedure. Moreover, it is possible to reduce the search space to the set $2^{\tilde{X}_{in}}$. In all our computational experiments we followed this path.

5 Computing Cluster Implementation of the Monte Carlo Algorithms for Constructing SAT Partitionings

The minimization algorithms suggested above were implemented as a parallel MPI-program PDSAT [37]. In PDSAT there is one leader process, all the other are follower processes (each process corresponds to 1 CPU core). The leader process in the PDSAT program selects point χ of the search space and creates random sample $\{\alpha_1, \dots, \alpha_N\}$. Follower processes solve corresponding SAT instances of the kind $C[\tilde{X}/\alpha_j]$, $j \in \{1, \dots, N\}$ (see section 4). The value of predictive function is always computed assuming that all the subproblems from the partitioning $\Delta(C, \tilde{X})$ will be processed by 1 CPU core.

We considered the inversion problem of the function (3): given the 114 bits of keystream we needed to find the secret key of length 64 bits, which produces this keystream (in accordance with the A5/1 algorithm). The PDSAT program was used to find partitionings with good time estimations for CNFs encoding this problem. The computational experiments were performed on the computing cluster Blackford ISDCT SB RAS¹ — in each experiment PDSAT was launched for 1 day using 16 Intel Xeon 5345 EM64T CPUs (64 cores in total). We used random samples of size $N = 10^4$.

On Figures 3, 4, 5 three decompositions sets are shown. We described the first set (further referred to as S1) in the paper [38] earlier. This set (of 31 variables) was constructed “manually” based on the analysis of algorithmic features of the A5/1 generator. The second one (S2), consisting of 31 variables, was found as a result of the minimization of $F(\cdot)$ by the simulated annealing algorithm (see section 4). The third set (S3), consisting of 32 variables, was found as a result of minimization of $F(\cdot)$ by the TS algorithm. In the table 1 the values of $F(\cdot)$ (in seconds) for all three sets are shown. Note that each of sets S_2 and S_3 was found for one 114 bit fragment of keystream that was generated according to the A5/1 algorithm for a randomly chosen 64-bit secret key.

Table 1. Decomposition sets for the A5/1 cryptanalysis problem and corresponding values of the predictive function.

Set	Power of set	$F(\cdot)$
S1	31	4.45140e+08
S2	31	4.78318e+08
S3	32	4.64428e+08

¹ <http://hpc.icc.ru/>

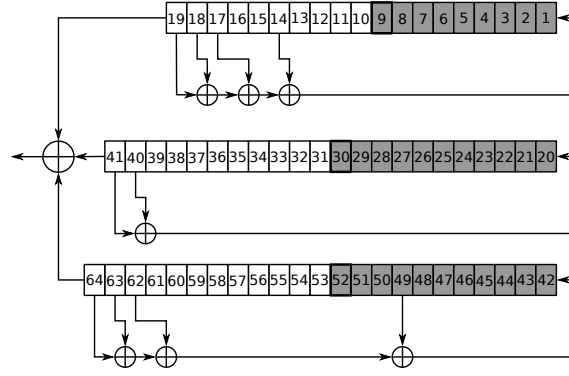


Fig. 3. Decomposition set S1

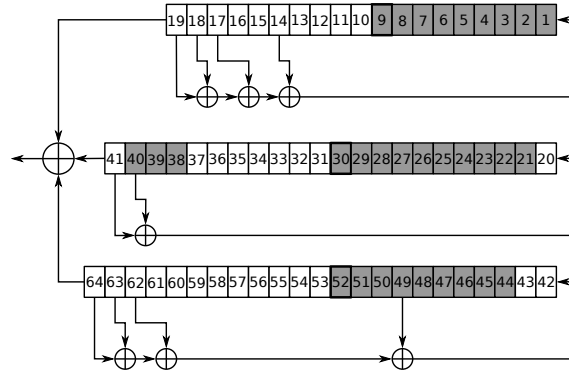


Fig. 4. Decomposition set S2

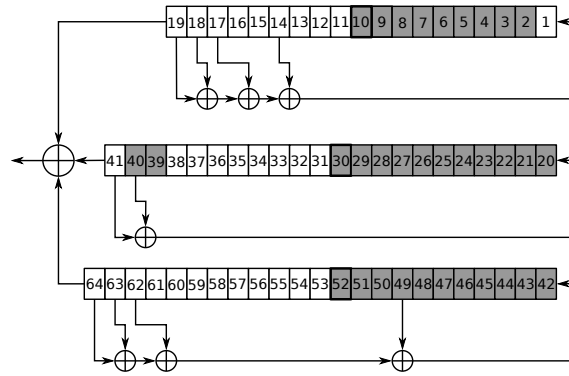


Fig. 5. Decomposition set S3

6 Cryptanalysis of the A5/1 Keystream Generator in the Volunteer Computing Project SAT@home

Volunteer computing [10] is a type of distributed computing which uses computational resources of PCs of private persons called volunteers. Each volunteer computing project is designed to solve one or several hard problems. When PC is connected to the project, all the calculations are performed automatically and do not inconvenience user since only idle resources of PC are used.

The first volunteer project was the GIMPS² project launched in 1996. Nowadays the most popular platform for organizing volunteer computing projects is BOINC [2] that was developed in Berkeley in 2002. Today there are about 70 active volunteer projects, the majority of them based on BOINC. The total performance of these projects exceeds 10 petaflops. Among the most important results obtained via volunteer computing are the discoveries of the largest prime number in the GIMPS project and of several radiopulsars in the Einstein@home³ project.

Volunteer computing project consists of the following basic parts: server daemons, database, web site and client application. Client application should have versions for the widespread computing platforms. One of the attractive features of volunteer computing is its low cost — to maintain a project one needs only a dedicated server working 24/7. Another advantage of this type of computing is that volunteer project can solve some particular hard problem for months or even years with good average performance.

SAT@home⁴ [35] is volunteer computing project aimed at solving hard combinatorial problems that can be effectively reduced to SAT. It was launched on September 29, 2011 by ISDCT SB RAS and IITP RAS. SAT@home was developed with the help of BOINC platform [2] and SZTAKI Desktop Grid package [23]. On February 7, 2012 SAT@home was added to the official list of BOINC projects⁵ with alpha status. Recently its status was improved to beta.

The SAT@home server uses a number of standard BOINC daemons responsible for sending and processing tasks (transitioner, feeder, scheduler, etc.). Such daemons as work generator, validator and assimilator were implemented taking into account the specificity of the project. The work generator decomposes the original SAT problem to subproblems according to the partitioning approach (see section 4). It creates 2 copies of each task in accordance with the concept of redundant calculations used in BOINC. The validator checks the correctness of the results, and the assimilator processes correct results. The client application is based on the SAT solver MINISAT [11].

The characteristics of the SAT@home project as of 12 of November 2014 are (according to BOINCstats⁶):

- 3113 active PCs (active PC in volunteer computing is a PC that sent at least one result in last 30 days) about 80% of them use Microsoft Windows OS;
- 1337 active users (active user is a user that has at least one active PC);

² <http://www.mersenne.org/>

³ <http://einstein.phys.uwm.edu/>

⁴ <http://sat.isa.ru/pdsat/>

⁵ <http://boinc.berkeley.edu/projects.php>

⁶ <http://boincstats.com/>

- versions of the client application for CPU: Windows x86, Windows x86-64, Linux x86, Linux x86-64;
- average real performance: 3.2 teraflops, maximal performance: 7.9 teraflops.

The dynamics of the real performance of SAT@home can be seen at the SAT@home performance page⁷.

The experiment consisting in solving 10 inversion problems of function $f_{A5/1}$ was held in SAT@home from December 2011 to May 2012. To construct the corresponding tests we used the known rainbow-tables for the A5/1 algorithm⁸. These tables provide about 88% probability of success when analyzing 8 bursts of keystream (i.e. 914 bits). We randomly generated 1000 problems and applied the rainbow-tables technique to analyze 8 bursts of keystream, generated by A5/1. Among these 1000 problems the rainbow-tables could not find the secret key for 125 problems. From these 125 problems we randomly chose 10 and in the computational experiments applied the SAT approach to the analysis of first bursts of the corresponding keystream fragments (114 bits). For each SAT instance we constructed the partitioning generated by the S1 decomposition set (see Figure 3) and processed it in the SAT@home project. All 10 instances constructed this way were successfully solved in SAT@home in about 5 months (the average performance of the project at that time was about 2 teraflops). After finding the satisfying assignment the processing of the corresponding partitioning was interrupted. On average to solve one SAT instance the SAT@home project had to process about 1 billion subproblems from the corresponding partitioning.

The second experiment on the cryptanalysis of A5/1 was launched in SAT@home in May 2014. It was done with the purpose of testing the decomposition sets found automatically with the help of the procedures described in section 4. In particular we took the decomposition set S3 (see Figure 5) that was constructed for the SAT instance encoding the cryptanalysis of one particular 114-bit A5/1 keystream fragment, and then used this set to solve the series of tests with 10 similar instances. On September 26 we successfully solved in SAT@home all 10 instances from the considered series. It should be noted that in all the experiments the time required to solve the problem agrees with the predictive function value computed for the S3 decomposition set.

The problem of cryptanalysis of the generator A5/1 is also interesting because the same keystream of an arbitrary length can be generated from different secret keys. This fact was noted by Jovan Golic in [15]. We refer to these situations as “collisions” using the evident analogy with the corresponding notion from the theory of hash functions. The approach presented in our paper allows us to solve the problem of finding all the collisions of the generator A5/1 for a given fragment of a keystream. Using the SAT@home project collisions for several SAT instance were found (see Table 2).

Our computational experiments clearly demonstrate that the proposed method of automatic search for decomposition sets makes it possible to construct SAT partitionings with the properties close to that of “reference” partitionings, i.e. partitionings constructed based on the analysis of algorithmic features of the considered cryptographic functions.

⁷ <http://sat.isa.ru/pdsat/performance.php>

⁸ <https://opensource.srlabs.de/projects/a51-decrypt>

Table 2. Original keys and collisions of the generator A5/1 (in hexadecimal format)

instance	original secret key	collision
1	F43FF04CD4F45660	7A1FF04CD4F45660
2	B95654F2242C6DF1	5CAB34F2242C6DF1
3	67685940B034EF78	B3B43940B034EF78
4	41038717BBA05E57	FB164EC44124398E

7 Related Works

In the book [3] the Boolean satisfiability problem is considered in the variety of aspects. The problems regarding the organization of the SAT solving process in distributed computing environments have been studied in [44,6,20,7,18,19]. The partitioning approach was analyzed in detail in [20,21,19]. A number of approaches to estimating the runtime for Constraint Satisfaction Problem and SAT can be found in [1,25,17,43].

The first example of the application of the SAT approach to the cryptanalysis problems is [30]. In [31,12,40,39] several estimations of time required to solve SAT for CNFs encoding the Bivium cipher cryptanalysis were proposed. The ideas underlying the cited works are close to the ones we use. The main distinction between our results and that of [31,12,40,39] is that the decomposition sets that were constructed in those papers were found based on either the structure of the encryption algorithm or randomly. Our main achievement consists in constructing automatic procedures to solve this problem.

The most effective in practice method of cryptanalysis of A5/1 is the Rainbow-method, partial description of which can be found on the A5/1 Cracking Project site⁹. In [16] the description of a number of techniques, used in the A5/1 Cracking Project to construct Rainbow tables, was presented. The cryptanalysis of A5/1 via Rainbow tables has the success rate of approximately 88% if one uses 8 bursts of keystream. The success rate of the Rainbow method if one has only 1 burst of keystream is about 24%. In all our computational experiments we analyzed the keystream fragment of size 114 bits, i.e. one burst. In [38] we described our first experience on the application of the SAT approach to A5/1 cryptanalysis in the specially constructed grid system BNB-Grid. The decomposition set used in [38] was constructed manually based on the peculiarities of the A5/1 algorithm.

The grid systems aimed at solving SAT are relatively rare. In [36] a desktop grid for solving SAT which used conflict clauses exchange via a peer-to-peer protocol was described. Apparently, [5] became the first paper about the use of a desktop grid based on the BOINC platform for solving SAT. Unfortunately, it did not evolve into a full-fledged volunteer computing project. As we noted above, the predecessor of the SAT@home was the BNB-Grid system [13,38], that was used to solve first large scale cryptographic problems in 2009.

⁹ <https://opensource.srlabs.de/projects/a51-decrypt>

8 Conclusions

In the present paper we describe the volunteer computing project SAT@home designed specifically for the purpose of long-term solving of hard SAT instances. We also propose the partitioning method for SAT instances encoding inversion problems of cryptographic functions. The applicability of proposed technologies is demonstrated on the problem of cryptanalysis of the A5/1 keystream generator. In the SAT@home project we managed to solve several cryptanalysis instances for this generator. In all the experiments we analyzed only the first burst (114 bits) of keystream.

9 Acknowledgements

Authors thank Stepan Kochemazov for numerous valuable comments that allowed us to significantly improve the quality of the paper, Alexey Ignatiev and Irina Bogachkova for constructive feedback and helpful discussions. This work was partly supported by Russian Foundation for Basic Research (grants 14-07-00403-a and 14-07-31172-mol-a) and by the President of Russian Federation grants for young scientists (SP-1855.2012.5 and SP-3667.2013.5).

References

1. Aloul, F.A., Sierawski, B.D., Sakallah, K.A.: Satometer: how much have we searched? *IEEE Trans. on CAD of Integrated Circuits and Systems* 22(8), 995–1004 (2003)
2. Anderson, D.P.: BOINC: A System for Public-Resource Computing and Storage. In: Buyya, R. (ed.) *GRID*. pp. 4–10. IEEE Computer Society (2004)
3. Biere, A., Heule, M., van Maaren, H., Walsh, T. (eds.): *Handbook of Satisfiability*, *Frontiers in Artificial Intelligence and Applications*, vol. 185. IOS Press (2009)
4. Biryukov, A., Shamir, A., Wagner, D.: Real Time Cryptanalysis of A5/1 on a PC. In: Schneier, B. (ed.) *FSE*. LNCS, vol. 1978, pp. 1–18. Springer (2000)
5. Black, M., Bard, G.: SAT Over BOINC: An Application-Independent Volunteer Grid Project. In: Jha, S., gentschen Felde, N., Buyya, R., Fedak, G. (eds.) *GRID*. pp. 226–227. IEEE (2011)
6. Blochinger, W., Sinz, C., Küchlin, W.: Parallel propositional satisfiability checking with distributed dynamic learning. *Parallel Computing* 29(7), 969–994 (2003)
7. Chrabakh, W., Wolski, R.: GridSAT: a system for solving satisfiability problems using a computational grid. *Parallel Computing* 32(9), 660–687 (2006)
8. Cook, S.A.: The complexity of theorem-proving procedures. In: Harrison, M.A., Banerji, R.B., Ullman, J.D. (eds.) *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, May 3-5, 1971, Shaker Heights, Ohio, USA. pp. 151–158. ACM (1971)
9. Davis, M., Logemann, G., Loveland, D.: A machine program for theorem-proving. *Commun. ACM* 5(7), 394–397 (Jul 1962)
10. Durrani, M.N., Shamsi, J.A.: Volunteer computing: requirements, challenges, and solutions. *J. Network and Computer Applications* 39, 369–380 (2014)
11. Eén, N., Sörensson, N.: An Extensible SAT-solver. In: Giunchiglia, E., Tacchella, A. (eds.) *SAT*. LNCS, vol. 2919, pp. 502–518. Springer (2003)
12. Eibach, T., Pilz, E., Völkel, G.: Attacking Bivium Using SAT Solvers. In: Büning, H.K., Zhao, X. (eds.) *SAT. Lecture Notes in Computer Science*, vol. 4996, pp. 63–76. Springer (2008)

13. Evtushenko, Y., Posypkin, M., Sigal, I.: A framework for parallel large-scale global optimization. *Computer Science - R&D* 23(3-4), 211–215 (2009)
14. Glover, F., Laguna, M.: *Tabu Search*. Kluwer Academic Publishers (1997)
15. Golic, J.D.: Cryptanalysis of alleged a5 stream cipher. In: *Proceedings of the 16th Annual International Conference on Theory and Application of Cryptographic Techniques*. pp. 239–255. EUROCRYPT’97, Springer-Verlag, Berlin, Heidelberg (1997)
16. Güneysu, T., Kasper, T., Novotný, M., Paar, C., Rupp, A.: Cryptanalysis with copacabana. *IEEE Trans. Comput.* 57(11), 1498–1513 (Nov 2008)
17. Haim, S., Walsh, T.: Online Estimation of SAT Solving Runtime. In: Büning, H.K., Zhao, X. (eds.) *SAT*. Lecture Notes in Computer Science, vol. 4996, pp. 133–138. Springer (2008)
18. Heule, M., Kullmann, O., Wieringa, S., Biere, A.: Cube and Conquer: Guiding CDCL SAT Solvers by Lookaheads. In: Eder, K., Lourenço, J., Shehory, O. (eds.) *Haifa Verification Conference*. LNCS, vol. 7261, pp. 50–65. Springer (2011)
19. Hyvärinen, A.E.J.: *Grid Based Propositional Satisfiability Solving*. Ph.D. thesis, Aalto University (2011)
20. Hyvärinen, A.E.J., Junttila, T.A., Niemelä, I.: A Distribution Method for Solving SAT in Grids. In: Biere, A., Gomes, C.P. (eds.) *SAT*. LNCS, vol. 4121, pp. 430–435. Springer (2006)
21. Hyvärinen, A.E.J., Junttila, T.A., Niemelä, I.: Partitioning SAT Instances for Distributed Solving. In: Fermüller, C.G., Voronkov, A. (eds.) *LPAR (Yogyakarta)*. LNCS, vol. 6397, pp. 372–386. Springer (2010)
22. Jarvisalo, M., Junttila, T.A.: Limitations of restricted branching in clause learning. *Constraints* 14(3), 325–356 (2009)
23. Kacsuk, P., Kovács, J., Farkas, Z., Marosi, A., Gombás, G., Balaton, Z.: Sztaki desktop grid (szdg): a flexible and scalable desktop grid system. *Journal of Grid Computing* 7(4), 439–461 (2009)
24. Kalos, M.H., Whitlock, P.A.: *Monte Carlo methods*. Wiley, New York, NY (1986)
25. Kilby, P., Slaney, J.K., Thiébaux, S., Walsh, T.: Estimating search tree size. In: *AAAI*. pp. 1014–1019. AAAI Press (2006)
26. King, J.C.: Symbolic execution and program testing. *Commun. ACM* 19(7), 385–394 (Jul 1976)
27. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *SCIENCE* 220(4598), 671–680 (1983)
28. Marques-Silva, J., Lynce, I., Malik, S.: Conflict-driven clause learning sat solvers. In: Biere et al. [3], pp. 131–153
29. Marques-Silva, J.P., Sakallah, K.A.: GRASP: A search algorithm for propositional satisfiability. *IEEE Trans. Computers* 48(5), 506–521 (1999)
30. Massacci, F., Marraro, L.: Logical Cryptanalysis as a SAT Problem. *J. Autom. Reasoning* 24(1/2), 165–203 (2000)
31. McDonald, C., Charnes, C., Pieprzyk, J.: Attacking Bivium with MiniSat. Tech. Rep. 2007/040, ECRYPT Stream Cipher Project (2007)
32. Metropolis, N., Ulam, S.: The Monte Carlo Method. *J. Amer. statistical assoc.* 44(247), 335–341 (1949)
33. Moskewicz, M.W., Madigan, C.F., Zhao, Y., Zhang, L., Malik, S.: Chaff: Engineering an efficient SAT solver. In: *Proceedings of the 38th Design Automation Conference, DAC 2001, Las Vegas, NV, USA, June 18-22, 2001*. pp. 530–535. ACM (2001)
34. Otpuschennikov, I., Semenov, A., Kochemazov, S.: Transalg: a tool for translating procedural descriptions of discrete functions to sat (tool paper). CoRR abs/1405.1544 (2014)
35. Posypkin, M., Semenov, A., Zaikin, O.: Using BOINC desktop grid to solve large scale SAT problems. *Computer Science Journal* 13(1), 25–34 (2012)
36. Schulz, S., Blochinger, W.: Parallel SAT Solving on Peer-to-Peer Desktop Grids. *J. Grid Comput.* 8(3), 443–471 (2010)

37. Semenov, A., Zaikin, O.: On estimating total time to solve SAT in distributed computing environments: Application to the SAT@home project. CoRR abs/1308.0761 (2013)
38. Semenov, A., Zaikin, O., Bessalov, D., Posypkin, M.: Parallel Logical Cryptanalysis of the Generator A5/1 in BNB-Grid System. In: Malyskin, V. (ed.) PaCT. LNCS, vol. 6873, pp. 473–483. Springer (2011)
39. Soos, M.: Grain of Salt - an Automated Way to Test Stream Ciphers through SAT Solvers. In: Tools'10: Proceedings of the Workshop on Tools for Cryptanalysis. pp. 131–144 (2010)
40. Soos, M., Nohl, K., Castelluccia, C.: Extending SAT Solvers to Cryptographic Problems. In: Kullmann, O. (ed.) SAT. LNCS, vol. 5584, pp. 244–257. Springer (2009)
41. Tseitin, G.S.: On the complexity of derivation in propositional calculus. In: Siekmann, J., Wrightson, G. (eds.) Automation of Reasoning 2: Classical Papers on Computational Logic 1967–1970, pp. 466–483. Springer, Berlin, Heidelberg (1983)
42. Williams, R., Gomes, C.P., Selman, B.: Backdoors to typical case complexity. In: Gottlob, G., Walsh, T. (eds.) IJCAI. pp. 1173–1178. Morgan Kaufmann (2003)
43. Xu, L., Hutter, F., Hoos, H.H., Leyton-Brown, K.: SATzilla: Portfolio-based Algorithm Selection for SAT. J. Artif. Intell. Res. (JAIR) 32, 565–606 (2008)
44. Zhang, H., Bonacina, M.P., Hsiang, J.: PSATO: a Distributed Propositional Prover and its Application to Quasigroup Problems. J. Symb. Comput. 21(4), 543–560 (1996)
45. Zhang, L., Madigan, C.F., Moskewicz, M.W., Malik, S.: Efficient conflict driven learning in boolean satisfiability solver. In: ICCAD. pp. 279–285 (2001)